

M. Caporin, P. Paruolo

Multivariate ARCH with spatial effects for stock sector and size

2005/13



UNIVERSITÀ DELL'INSUBRIA
FACOLTÀ DI ECONOMIA

<http://eco.uninsubria.it>

In questi quaderni vengono pubblicati i lavori dei docenti della Facoltà di Economia dell'Università dell'Insubria. La pubblicazione di contributi di altri studiosi, che abbiano un rapporto didattico o scientifico stabile con la Facoltà, può essere proposta da un professore della Facoltà, dopo che il contributo sia stato discusso pubblicamente. Il nome del proponente è riportato in nota all'articolo. I punti di vista espressi nei quaderni della Facoltà di Economia riflettono unicamente le opinioni degli autori, e non rispecchiano necessariamente quelli della Facoltà di Economia dell'Università dell'Insubria.

These Working papers collect the work of the Faculty of Economics of the University of Insubria. The publication of work by other Authors can be proposed by a member of the Faculty, provided that the paper has been presented in public. The name of the proposer is reported in a footnote. The views expressed in the Working papers reflect the opinions of the Authors only, and not necessarily the ones of the Economics Faculty of the University of Insubria.

© Copyright M. Caporin, P. Paruolo
Printed in Italy in December 2005
Università degli Studi dell'Insubria
Via Monte Generoso, 71, 21100 Varese, Italy

All rights reserved. No part of this paper may be reproduced in any form without permission of the Author.

Multivariate ARCH with spatial effects for stock sector and size*

Massimiliano Caporin
Università di Padova, Italy

Paolo Paruolo[†]
Università dell'Insubria, Varese, Italy

December 6, 2005

Abstract

This paper applies a new spatial approach for the specification of multivariate GARCH models, called Spatial Effects in ARCH, SEARCH. We consider spatial dependence associated with industrial sectors and capitalization size. This parametrization extends current feasible specifications for large scale GARCH models, keeping the numbers of parameters linear as a function of the number of assets. An application to daily returns on 150 stocks from the NYSE for the period January 1994 to June 2001 shows the benefits of the present specification when compared to alternative specifications.

Keywords: Spatial models, GARCH, Volatility, Large scale models, Portfolio allocation.

JEL code: C32, C51, C52.

*This paper was presented at the JAE conference 'Changing structures in International and Financial Markets and the effects on financial decision making', Venice (Italy) June 2-3, 2005. A previous version was also presented at the 'First Italian conference on econometrics and applied economics', Venice (Italy) January 24-25, 2005. We thank participants of both conferences for useful comments. Partial financial support from the Italian Ministry MIUR is gratefully acknowledged by both authors. The second author also acknowledges partial financial support from University of Insubria FAR grants 2004 and 2005.

[†] Corresponding author. Address: Paolo Paruolo, Via Monte Generoso 71, 21100, Varese, Italy. E-mail: paolo.paruolo@uninsubria.it. Tel. +39 0332 395500, fax +39 0332 395509.

Contents

1	Introduction	3
2	A general class of GARCH models	4
2.1	Conditional mean	4
2.2	Conditional variance	4
3	Spatial covariance structures	5
3.1	Asset classification	6
3.2	Spatial matrices	6
3.3	Matrices with spatial structure	6
3.4	S specification	7
3.5	Restricted S specification	8
3.6	EA dynamics	9
3.7	R specification	10
4	Estimation results	11
4.1	The data	11
4.2	Conditional mean	12
4.3	Estimation strategy for the conditional variance	12
4.4	S estimates	13
4.5	EA estimates	14
4.6	R estimates	14
5	Model evaluation	15
5.1	Residual analysis	15
5.2	Implied portfolio management	16
5.3	Volatility performance indicators	17
6	Conclusions	18
A	Appendix: List of assets by sector and size	19
B	Graphs and tables	21

1 Introduction

As it is well recognized in applied econometrics, unrestricted multivariate GARCH models lack parameter parsimony, at least for data-sets with a large cross-section dimension; see e.g. the review paper by Bauwens et al. (2003). In general multivariate GARCH specifications, the number of parameters – i.e. the model dimension – is in fact proportional to the fourth power of the number of assets, see Engle and Kroner (1995). This feature has limited the application of unrestricted GARCH models to systems of limited dimensions, i.e. typically with up to a handful of assets.

In this paper we consider a spatial approach in the specification of a GARCH models for stock returns. This model class, called SEARCH, was proposed in Caporin and Paruolo (2005), where properties of the model are discussed. In the present paper we apply this approach to 150 assets from the New York Stock Exchange, NYSE. This represents a large cross section of stock returns, at least for the estimation of GARCH processes.

The SEARCH model is compared with other feasible alternative specifications, such as the constant conditional correlation (CC) model of Bollerslev (1990) and the orthogonal GARCH of Alexander (2001), OG. It is found that the SEARCH specification favorably compares with alternatives, although it has a much smaller number of parameters.

The SEARCH model can be interpreted as an explicit factor model. Unlike in (so-called ‘exploratory’) factor analysis where factors are economically ex-ante unidentified, SEARCH models incorporate a factor structure built on economically-interpretable primary characteristics of assets, such as sector and size.

Thanks to the parametric nature of SEARCH, the econometrician can test if various levels of the economic factors have the same effect on volatility. This allows to investigate if and how sector and size play a role in the multivariate volatility structure; in case they do, the model directly provides estimates of their relative importance.

Spatial econometrics has a long history, see e.g. Anselin (1988) and Case (1991) and reference therein. Recent references include Conley and Topa (2002), Giacomini and Granger (2004), Pesaran et al. (2004), Baltagi et al. (2003). However, applications of ideas from spatial statistics to multivariate GARCH modeling are unknown to the authors; the present paper hence explores a novel approach in GARCH specification.

The aim of this paper is twofold. First, this paper wishes to study the empirical applicability and to illustrate the interpretation of SEARCH models on a large cross sections of assets. Secondly the paper presents a comparison of the performance of SEARCH models with alternatives in current use. Model comparison is here based on indicators derived from financial asset-allocation.

We find that SEARCH models can be fitted to large cross sections with a limited computational effort. The estimates reveal similarities and differences of volatility structures, within and across sectors and capitalization sizes. The possibility to explicitly interpret coefficients opens the door to several specification strategies based on sequences of nested models of the SEARCH class.

The performance comparison shows that, possibly thanks to the fact that SEARCH

models are parsimoniously parameterized, SEARCH models outperform competitors on the given asset cross-section. Although this finding cannot be generalized to other data-sets, it suggests that spatial ideas may provide a sensible approach to parsimonious volatility modeling.

The rest of the paper is organized as follows. Section 2 presents the general class of GARCH processes, which nests the CC, OG and SEARCH classes. Section 3 defines the SEARCH models as the spatially restricted version of this general class of models. Estimation results are reported in Sections 4, while model evaluation is reported in Section 5. Section 6 concludes.

In the paper we use the following notation: $a := b$ or $b =: a$ indicates that a is defined by b ; vec denotes the column stacking operator; $\text{vecd}(A)$ indicates a column vector containing the diagonal elements of a matrix A ; for any vector $a := (a_1 : \dots : a_n)'$, $\text{diag}(a) := \text{diag}(a_1, \dots, a_n)$ is a diagonal matrix with diagonal elements equal to a ; $\text{dg}(A)$ is a matrix with off-diagonal elements equal to 0 and diagonal elements equal to the ones on the diagonal of A ; ι_n indicates a $n \times 1$ vector of ones and I_n is the identity matrix of order n .

2 A general class of GARCH models

In this section we define a general class of GARCH models. This class includes OG, CC and SEARCH models as special cases. In Subsection 2.1 we introduce the conditional mean parameters, while in Subsection 2.2 we define the model for the conditional variance.

2.1 Conditional mean

Consider an $n_x \times 1$ dimensional vector time series $\{x_t\}_{t \in \mathbb{N}}$ and the associated filtration $\mathcal{I}_t := \sigma(x_{t-q}, q \geq 0)$. Let also $x_t := (y_t' : z_t')'$ be partitioned into a $n \times 1$ subvector of variables of interest y_t and a $n_z \times 1$ subvector of other information variables z_t . We assume that x_t has finite second moments conditional on \mathcal{I}_{t-1} . Indicate by $\mathbb{E}_{t-1}(\cdot) := \mathbb{E}(\cdot | \mathcal{I}_{t-1})$ the conditional expectation operator, and let $\mu_t := \mathbb{E}_{t-1}(y_t)$ be the conditional mean of y_t .

We consider a parametric model for the conditional mean μ_t , taken for simplicity to be linear, $\mu_t = \mu w_t$ where $w_t := (y_{t-1}' : \dots : y_{t-p_\mu}'', z_t')'$ is a $n_w \times 1$ vector and μ is a matrix of coefficients:

$$y_t =: \mathbb{E}_{t-1}(y_t) + \varepsilon_t = \mu w_t + \varepsilon_t, \quad (1)$$

Here $\varepsilon_t := y_t - \mu_t$ is the $n \times 1$ vector of deviations from the conditional mean.

2.2 Conditional variance

We assume that the cross section dependence of ε_t due to asset proximity can be summarized in a linear relation of the form

$$\varepsilon_t = S\varepsilon_t + \eta_t, \quad (2)$$

where η_t is a $n \times 1$ vector of random variables with $\mathbb{E}_{t-1}(\eta_t) = 0$ and $\mathbb{V}_{t-1}(\eta_t) = \mathbb{E}_{t-1}(\eta_t \eta_t') =: \Gamma_t$. Eq. (2) is similar to standard specifications in the structural VAR literature, see e.g. Amisano and Giannini (1997). The matrix $I - S$ plays the role of the matrix of covariance eigenvectors in OG models, see Alexander (2001). SEARCH models, defined in Section 3 below, define S to be of spatial structure, in the sense defined in Subsection 3.3. In this case eq. (2) defines a spatial autoregression (SAR) process, see Cressie (1993) eq. (6.3.8) and reference therein.

We indicate by Σ_t the conditional variance covariance matrix of ε_t , $\Sigma_t := \mathbb{V}_{t-1}(\varepsilon_t) = \mathbb{E}_{t-1}(\varepsilon_t \varepsilon_t')$. A direct consequence of the assumption (2) is that

$$\Sigma_t := \mathbb{V}_{t-1}(\varepsilon_t) = (I - S)^{-1} \Gamma_t (I - S')^{-1}.$$

provided $I - S$ is invertible, which we assume in the following. Similarly to the CC class, we assume that the errors η_t in (2) have constant correlations, i.e.

$$\Gamma_t = D_t R D_t \quad (3)$$

where D_t is diagonal and $D_t^2 = \text{diag}(h_t) := \text{diag}(h_{1t}, \dots, h_{nt})$ with (possibly) time varying h_{it} and R is a time-invariant correlation matrix of the standardized innovations $\psi_t := D_t^{-1}(I - S)\varepsilon_t = D_t^{-1}\eta_t$, $\mathbb{V}_{t-1}(\psi_t) = \mathbb{E}_{t-1}(\psi_t \psi_t') = R$.

We next consider GARCH(p_E, p_A) dynamics for $h_t := \text{vecd}(D_t^2)$ driven by $e_t := (\eta_{1,t}^2, \dots, \eta_{n,t}^2)'$

$$h_t = \zeta + E(L) h_{t-1} + A(L) e_{t-1}, \quad (4)$$

where by definitions above $h_t = \mathbb{E}_{t-1}(e_t)$, so that $e_t - h_t$ is an innovation with respect to \mathcal{I}_{t-1} . This allows to calculate multi-step ahead predictions for the conditional variances using recursions as in standard multivariate GARCH models. Here $E(L) := \sum_{l=1}^{p_E} E_l L^{l-1}$ and $A(L) := \sum_{l=1}^{p_A} A_l L^{l-1}$, E_l, A_l are $n \times n$ matrices, ζ is a n -dimensional vector.

Model (2)-(4) nests both the OG and the CC classes, which can be obtained for special choices of the matrices S, E_l, A_l, R . The general model (2)-(4), see Caporin and Paruolo (2005) could be further extended, in order to incorporate leverage effect and various functional transformation on the conditional variances and/or on the innovations; for simplicity this is not further explored in the present paper.

The general model (2)-(4) has dimension that is quadratic in the number of assets. In fact the unrestricted matrices S, E_l, A_l, R contain $O(n^2)$ parameters. Further restrictions are hence needed in order to render estimation of the model feasible on large cross sections.

In the following we indicate with θ the $v \times 1$ vector of parameters. Parameters within θ are partitioned into $\theta = (\theta'_\mu, \theta'_S, \theta'_{EA}, \theta'_R)'$ where $\theta_S, \theta_{EA}, \theta_R$ are the subvectors of parameters in the S, EA and R specification. The dimension of the subvectors $\theta_S, \theta_{EA}, \theta_R$ is indicated as v_S, v_{EA}, v_R respectively, with $v = v_\mu + v_S + v_{EA} + v_R$. The θ_{EA} parameters are further partitioned into $\zeta, \theta_E, \theta_A$ in an obvious notation.

3 Spatial covariance structures

The SEARCH class of models is characterized as the submodel of (2)-(4) with the property that the matrices S, E_l, A_l, R (SEAR) which define the conditional het-

eroscedasticity (CH) features of the model, have *spatial structure* in the sense defined in Subsection 3.3.

In this section we discuss the notion of spatial dependence. An asset classification based on stocks sectors and size is first presented in Subsection 3.1; we next define an implied notion of proximity in Subsection 3.2. The definition of matrices with (generalized) spatial structure is provided in Subsection 3.3. In Subsections 3.4, 3.5, 3.6 and 3.7 we discuss the unrestricted and restricted specifications for the S matrix, the GARCH dynamics, and the correlation matrix R , respectively.

3.1 Asset classification

Consider the vector of returns $y_t := (y_{1t}, \dots, y_{nt})'$, where y_{qt} is the return on asset q , where q belongs to the set $\mathcal{C}_\cdot := \{1, 2, \dots, n\}$ of the first n integers. We assume that there exist a time-invariant classification of the assets as specified in Table 1, where

$$\mathcal{C}_{ij} := \{q \in \mathcal{C}_\cdot : \text{stock } q \text{ belongs to sector } i \text{ and capitalization size } j\} \quad (5)$$

indicates the set of labels of the stocks that belong to industrial sector i and capitalization class j . Define also the aggregated label sets for each sector $\mathcal{C}_i := \cup_{j=1}^\ell \mathcal{C}_{ij}$, for each capitalization class $\mathcal{C}_j := \cup_{i=1}^k \mathcal{C}_{ij}$ and for the whole market $\mathcal{C}_\cdot = \cup_{i=1}^k \cup_{j=1}^\ell \mathcal{C}_{ij} = \cup_{i=1}^k \mathcal{C}_i = \cup_{j=1}^\ell \mathcal{C}_j$. Classes \mathcal{C}_{ij} are assumed to be exhaustive and mutually exclusive. The spatial covariance specification considered in this paper is based on such a classification.

[Table 1]

3.2 Spatial matrices

A matrix $W^\circ = (w_{ij}^\circ)$ is called a spatial weight matrix, or simply spatial matrix, if it has real entries contained in the closed interval $[0, 1]$, and diagonal elements equal to 0, i.e. $0 \leq w_{ij}^\circ \leq 1$, $\text{vecd}(W^\circ) = 0$. A matrix W° is called a spatial lag matrix if W° is a spatial matrix and $w_{ij}^\circ > 0$ only for first-order neighbors j of unit i . If the row-sums equal 1, $W^\circ \iota_n = \iota_n$, then the spatial matrix is called ‘normalized’.

In the following we use different spatial matrices W° for different matrices in the set S , E_l , A_l , R . In this way we model between-classes linkages through S and intra-class linkages via E_l , A_l , R . Table 1 suggests to define first-order neighbors as the stocks with the same combination of sector and size, or just the same sector, or just the same size class.

In the next subsection we specify the connection between S , E_l , A_l , R and the W° matrices.

3.3 Matrices with spatial structure

In this subsection we define matrices of (generalized) spatial structure. Let C indicate any of the S , E_l , A_l , R matrices. We say that C has spatial structure if it can be written as a linear combination of some spatial weight matrices W_q° , $q = 0, 1, \dots, m$, i.e. if

$$C = c_0 I_n + \sum_{q=1}^m c_q W_q^\circ \quad (6)$$

for $m \geq 1$ and $W_1^\circ, \dots, W_m^\circ$ pre-defined spatial weight matrices. When c_0, c_1, \dots, c_m are real scalars, we call eq. (6) a *scalar* specification.

In order to allow for heterogeneity across units in the spatial specification, we enlarge this definition allowing c_q in (6) to be diagonal matrices, $c_q := \text{diag}(c_{q,1} : \dots : c_{q,n})'$, when n is the dimension of W_q° . In this case we say that matrix C in (6) has a *generalized* spatial structure, or, equivalently that C has a *diagonal* (spatial) specification.

When one or more rows of W_q° have all 0 entries, we define the corresponding entries $c_{q,i}$ to be equal to zero, i.e. $J'W_q^\circ = 0$ implies $c_q J = 0$ for any full-column-rank matrix J . When $c_{q,1} = \dots = c_{q,n}$ the diagonal specification reduces to the scalar one, i.e. a diagonal specification nests a scalar one.

In (6) the elements $c_{q,i}$ represent (possibly different) loadings on (the level of) a factor, as detailed below for the matrix S . For simplicity the number of elements m in the sum is assumed to be the same for all rows i , but this format can easily accommodate different number of terms in the sum by setting to zero appropriate elements $c_{q,i}$.

3.4 S specification

S is restricted to be of spatial nature with all diagonal elements equal to zero, in order to exclude simultaneous effects from ε_{qt} to itself, i.e.

$$S = \sum_{q=1}^{m_S} s_q W_q, \quad (7)$$

for an appropriate definition of the W_q spatial matrices. The elements s_i must guarantee that $I - S$ is invertible, but are not otherwise constrained.

We next define the spatial lag matrices W_q in (7). Let $W(\mathcal{C}_{ij}) := (w_{ql})$, where $w_{ql} := (n_{ij} - 1)^{-1} 1(q \in \mathcal{C}_{ij}) (1(l \in \mathcal{C}_{ij}) - \delta_{ql})$ be the normalized spatial matrix that defines as neighbors of the q -th unit the units that belong to the same class \mathcal{C}_{ij} ; here $n_{ij} := \#(\mathcal{C}_{ij})$ is the cardinality of the set \mathcal{C}_{ij} and $1(\cdot)$ is the indicator function. We choose the W_q matrices for S in (7) as $W_{ij} := W(\mathcal{C}_{ij})$, $i = \cdot, 1, \dots, k$, $j = \cdot, 1, \dots, \ell$, i.e.

$$S = s_{..} W_{..} + \sum_{i=1}^k s_{i.} W_{i.} + \sum_{j=1}^{\ell} s_{.j} W_{.j} + \sum_{i=1}^k \sum_{j=1}^{\ell} s_{ij} W_{ij}. \quad (8)$$

The following remarks are in order.

1. Market-wide spatial matrix

Consider $W_{..} := W(\mathcal{C}_{..}) = (n - 1)^{-1} (\iota_n \iota_n' - I_n)$. This matrix specifies the dependence of ε_{qt} on the average of all remaining ε_{qt} . In fact let $\bar{\varepsilon}_t^{(-q)} := (n - 1)^{-1} \sum_{l \in \mathcal{C}_{..}, l \neq q} \varepsilon_{lt}$ be the leave-one-out arithmetic average of all the ε_{lt} , and recall that $\varepsilon_t = S\varepsilon_t + \eta_t$. Then the term $s_{..} W_{..} \varepsilon_t$ on the r.h.s. can be

expressed as follows:

$$s_{..}W_{..}\varepsilon_t = s_{..}\frac{1}{n-1}(\iota_n\iota_n'\varepsilon_t - \varepsilon_t) = s_{..}\begin{pmatrix} \bar{\varepsilon}_t^{(-1)} \\ \bar{\varepsilon}_t^{(-2)} \\ \vdots \\ \bar{\varepsilon}_t^{(-n)} \end{pmatrix}. \quad (9)$$

Hence (9) describes the dependence of ε_{qt} on the average of all the remaining ε s. $s_{..}W_{..}$ hence captures the effect of a market-wide factor that affects all stocks simultaneously with equal loading if $s_{..}$ is a scalar, or with different loadings for each asset if $s_{..}$ is a diagonal matrix.

2. Industrial sector i spatial matrix

Consider $W_{i.} := W(\mathcal{C}_{i.})$ for $i = 1, \dots, k$. This matrix expresses the dependence of ε_{qt} on the arithmetic average of the contemporaneous ε_{qt} that belong to the same sector i . A representation similar to (9) holds in terms of the leave-one-out arithmetic average of the ε_{qt} for $q \in \mathcal{C}_{i.}$, with positive terms for the rows $q \in \mathcal{C}_{i.}$, and zero otherwise. Hence $W_{i.}$ is associated with a sector-specific factor for sector i ; we say that $W_{i.}\varepsilon_t$ represents the i -th *level* of the factor ‘sector’, $i = 1, \dots, k$.

3. Capitalization class j spatial matrix

Consider $W_{.j} := W(\mathcal{C}_{.j})$, $j = 1, \dots, \ell$. This matrix refers to all stocks in the same capitalization size, and it is thus associated with a size-specific factor for size j . We say that $W_{.j}\varepsilon_t$ represents the j -th *level* of the factor ‘size’, $j = 1, \dots, \ell$.

4. Interaction spatial matrices

Consider $W_{ij} := W(\mathcal{C}_{ij})$, $i = 1, \dots, k$, $j = 1, \dots, \ell$. This matrix expresses the dependence of ε_{qt} on the ε_{qt} of the same cell, i.e. that have simultaneously the same sector and size. Here W_{ij} is associated with a sector-and-size-specific factor; we say that $W_{ij}\varepsilon_t$ represents the (i,j) -th *level* of the factor (sector-and-size) ‘interaction’, $i = 1, \dots, k$, $j = 1, \dots, \ell$.

3.5 Restricted S specification

Eq. (8) nests several specifications with different degrees of parameter parsimony. One may e.g. restrict (8) in the following way:

$$s_{i.} = s_{0.} \text{ for all } i, \quad (10)$$

$$s_{.j} = s_{.0} \text{ for all } j, \quad (11)$$

$$s_{ij} = s_{00} \text{ for all } i \text{ and } j. \quad (12)$$

These restrictions imply homogeneous coefficients for different levels of the factors. When (10), (11), (12) hold, one can define the following aggregate spatial matrices

for all industrial sectors $W_{0\cdot}$, capitalization levels $W_{\cdot 0}$ and interactions W_{00} :

$$W_{0\cdot} := \sum_{i=1}^k W_{i\cdot} \quad W_{\cdot 0} := \sum_{j=1}^{\ell} W_{\cdot j} \quad W_{00} := \sum_{i=1}^k \sum_{j=1}^{\ell} W_{ij}. \quad (13)$$

With these definition, the restricted model for S can be written as

$$S = s_{\cdot\cdot} W_{\cdot\cdot} + s_{0\cdot} W_{0\cdot} + s_{\cdot 0} W_{\cdot 0} + s_{00} W_{00}. \quad (14)$$

We call (8) the unrestricted and (14) the restricted specifications. A prefix R is used to signal ‘restricted’. Each of them may have scalar or diagonal s_{ij} coefficients. We label the diagonal specifications for S as HET, for ‘heterogeneous’ responses of different assets; the scalar specifications for S are instead labeled HOM for ‘homogeneous’. One hence obtains 4 polar combinations, models HET, RHET, HOM, RHOM, nested as follows

$$\begin{array}{ccc} \text{HET} & \supset & \text{RHET} \\ \cup & & \cup \\ \text{HOM} & \supset & \text{RHOM} \end{array} \quad (15)$$

where HET, RHET, HOM, RHOM here represent the corresponding parameter sets. Note that HOM and RHET are non-nested, and that several intermediate model exists, where some of the s_{ij} coefficients are scalar and some other diagonal, or where only a subset of the restrictions (10), (11), (12) holds. We here note that, given the nesting structure in (15), many general-to-specific specification searches can be envisaged.

3.6 EA dynamics

We next consider the specification of the matrix polynomials $E(L) := \sum_{l=1}^{p_E} E_l L^{l-1}$ and $A(L) := \sum_{l=1}^{p_A} A_l L^{l-1}$ in (4), i.e. $h_t = \zeta + E(L) h_{t-1} + A(L) e_{t-1}$. The main interaction *among* classes \mathcal{C}_{ij} , see (5), is addressed via the S -specification. The ARCH dynamics in (4) is hence assumed to reflect a possible spatial dependence *within* each class \mathcal{C}_{ij} .

Without loss of generality, assume that assets are ordered as follows: the first n_{11} assets belong to the class \mathcal{C}_{11} , the next n_{12} belong to the class \mathcal{C}_{12} , and so forth, proceeding row-wise with respect to Table 1, where n_{ij} indicates the number of assets in class \mathcal{C}_{ij} .

Consider the partition of e_t into subvectors $e_t := (e'_{11,t}, e'_{12,t}, \dots, e'_{1\ell,t}, e'_{21,t}, \dots, e'_{k\ell,t})'$, where $e_{ij,t}$ is the subvector of e_t corresponding to the class \mathcal{C}_{ij} . Partition also η_t , ζ and h_t conformably. We assume that the $n \times n$ matrices E_l , A_l in $E(L) := \sum_{l=1}^p E_l L^{l-1}$ and $A(L) := \sum_{l=1}^q A_l L^{l-1}$ are block diagonal, where blocks are conformable with the partition of e_t . We indicate the blocks of E_l and A_l corresponding to $e_{ij,t-1}$ as $E_{ij,l}$ and $A_{ij,l}$ respectively; similarly we indicate the corresponding blocks of $E(L)$, $A(L)$ as $E_{ij}(L)$, $A_{ij}(L)$.

Note that the standard requirements for second-order stationarity of GARCH processes apply, i.e. the roots of $|I - E(z) - A(z)| = 0$ must be outside the unit circle. By the block-diagonal assumption of $E(L)$ and $A(L)$, this corresponds to the requirement that the roots of $|I_{n_{ij}} - E_{ij}(z) - A_{ij}(z)| = 0$ are outside the unit circle.

In order to define a spatial structure on the $E_{ij,l}$, $A_{ij,l}$ matrices, we assume that assets are ordered within each class \mathcal{C}_{ij} according to a proximity criterion. Many criteria may be considered, such as the ones based on earnings before income and taxes, dividend/price ratios, dividend/earning ratios etc. In absence of specific information, a single ordering criterion within each class \mathcal{C}_{ij} may be given by capitalization size; this is the case we adopt in the following.

In Caporin and Paruolo (2005) we consider two main EA-specifications; the first one is of spatial nature, the second one allows for a factor structure. Both can be seen as special cases of the spatial one, which we report here. The intercept ζ_{ij} is a linear function of some underlying vector of parameters $\theta_{c_{ij}}$, $\zeta_{ij} = \mathcal{F}_{ij}\theta_{c_{ij}}$ where \mathcal{F}_{ij} is a known square matrix of order n_{ij} . In the standard spatial specification, $\mathcal{F}_{ij} = I_{n_{ij}}$.

We assume that $E_{ij,l}$ and $A_{ij,l}$ have a spatial structure of the following form:

$$E_{ij,l} := \sum_{q=0}^{m_{E_{ij,l}}} \beta_{ij,lq} W_{ij,q}^* \quad A_{ij,l} := \sum_{q=0}^{m_{A_{ij,l}}} \alpha_{ij,lq} W_{ij,q}^*, \quad (16)$$

where $\beta_{ij,lq}$ and $\alpha_{ij,lq}$ are scalars or diagonal matrices and $W_{ij,0}^* := I_{n_{ij}}$. For $q > 0$ the matrices $W_{ij,q}^* := \text{diag}(W_{ij,q}^\circ \iota_{n_{ij}})^{-1} W_{ij,q}^\circ$ are $n_{ij} \times n_{ij}$ normalized spatial weight matrices that reflect proximity according to the intra-class ordering criterion.

Two possible choices for the spatial matrices $W_{ij,q}^\circ$ are the following

$$W_{ij,q}^\circ := U_{ij}^{q'} \quad (17)$$

$$\text{or} \quad W_{ij,q}^\circ := U_{ij}^q + U_{ij}^{q'} \quad \text{where} \quad U_{ij} := \begin{pmatrix} 0 & 0 \\ I_{n_{ij}-1} & 0 \end{pmatrix}. \quad (18)$$

For scalar $\alpha_{ij,lq}$ and $\beta_{ij,lq}$ parameters, this choice of $W_{ij,q}^\circ$ implies that $E_{ij,l}$ and $A_{ij,l}$ are Toeplitz matrices.

These specifications have the following interpretation: the spatial structure within each block relates each stock within $e_{ij,t}$ with the preceding one in the list; this is the case for specification (17), which implies an upper triangular system.

Alternatively, consider specification (18); each stock within $e_{ij,t}$ is related to the one preceding and the one following it in the list, which implies a symmetric Toeplitz matrix for the scalar specification. The lower triangular system can be obtained from (17) simply reversing the order within the block, and hence it is not treated separately.

The conditions for positive definiteness of the conditional variance matrix require the EA dynamics to deliver always positive definite conditional variances h_t . A sufficient condition for this is that the ζ , α and β parameters are positive.

3.7 R specification

Consider next the correlation matrix R , a positive definite, symmetric matrix with ones on the main diagonal. Again employing the same order of the blocks as in the previous subsection, we assume that R is block diagonal with diagonal blocks R_{ij} , where the subscripts ij refer to the class \mathcal{C}_{ij} . R_{ij} describes intra-class correlations

within class \mathcal{C}_{ij} . Within each diagonal block R_{ij} , we consider the following spatial specification:

$$R_{ij} = I_{n_{ij}} + \sum_{q=1}^{m_{R_{ij}}} \rho_{ij,q} W_{ij,q}^* \quad (19)$$

where $W_{ij,q}^\circ := U_{ij}^q + U_{ij}^{q'}$, $W_{ij,q}^* := \text{diag}(W_{ij,q}^\circ \iota_{n_{ij}})^{-1} W_{ij,q}^\circ$ and U_{ij} is defined in (18), $m_{R_{ij}} \leq n_{ij} - 1$. Note that the spatial nature of the R specification matches the one for the EA dynamics.

For example for $n_{ij} = 4$ one has

$$R_{ij} = \begin{pmatrix} 1 & \rho_{ij,1} & \rho_{ij,2} & \rho_{ij,3} \\ \rho_{ij,1} & 1 & \rho_{ij,1} & \rho_{ij,2} \\ \rho_{ij,2} & \rho_{ij,1} & 1 & \rho_{ij,1} \\ \rho_{ij,3} & \rho_{ij,2} & \rho_{ij,1} & 1 \end{pmatrix}.$$

The number of parameters in the R -scalar specification (19) is $n_{ij} - 1$ for each class \mathcal{C}_{ij} , for a total number of parameters equal to $v_R = \sum_{i,j} (n_{ij} - 1) = n - k\ell$. Note that the R -diagonal specification does not make sense, given the symmetry of each block R_{ij} . We hence consider only scalar R specifications.

A special case of (19) is given by the spatial autoregression of order one, where $\rho_{ij,q} = \rho_{ij}^q$ for some scalar ρ_{ij} . In this case the number of parameters reduces to 1 for each class \mathcal{C}_{ij} , for a total number of parameters equal to $v_R = k\ell$.

Another special case of interest is given by the restriction $\rho_{ij,q} = \rho_{ij}$ for some scalar ρ_{ij} and all q . This corresponds to the spatial classification of all the assets within class \mathcal{C}_{ij} as first-order neighbors. Also for this special case, there is only one parameters per class, for a total number of parameters equal to $v_R = k\ell$.

A number of possible extensions and variations can be considered on the basic scheme proposed here, see Caporin and Paruolo (2005), who also discuss the relation of the SEARCH model with other GARCH specifications. They show that all the SEARCH specifications are linear in n , the number of assets.

We next present the empirical results.

4 Estimation results

In this section we present the data and the estimation results on 150 stocks listed at the New York Stock exchange. In Subsection 4.1 we describe the data-set. Estimation of the conditional mean is reported in Subsection 4.2 while Subsection 4.3 reports the estimation strategy used in the estimation of the conditional variance.

In particular, estimation results for S are reported in Subsection 4.4, the ones for the GARCH parameters ζ , E , A are reported in Subsection 4.5, while estimated correlations in R are reported in Subsection 4.6.

4.1 The data

We selected 150 assets from the New York Stock Exchange. The data corrected for dividends and stock splits. The list of assets is reported in the Appendix.

The assets were chosen in order to satisfy the following criteria: (i) daily data is continuously available since January 1994; (ii) each asset belongs to one of the Standard & Poors indices based on firms capitalization (SP500 for large firms, SP400 for mid-cap firms and SP600 for small-cap firms, called large, L, medium, M, and small, S, in the following); (iii) 50 assets are chosen for each capitalization class; (iv) the assets belong to ten sectors: Basic Materials (BM), Capital Goods (CG), Consumer Cyclical Goods (CCG), Consumer Non-cyclical Goods (CNG), Energy (EN), Financial (FI), Healthcare (HE), Services (SE), Technology (TE), Utilities (UT); (v) for each sector we choose 15 assets, 5 in each of the three capitalization classes L, M, S.

In the empirical analysis we restricted the sample to the period January 1994 - June 2001, in order to exclude the technology market bubble. The sample consists of 1890 daily observations; the last 20% of the data (January 2000 to June 2001, starting at observation number 1514) was used for model evaluation, while the remaining part of the data was included in the estimation sample. The last 500 observation of the estimation sample were used for back-fitting evaluation.

The analysis was then performed on the log-returns $r_t = \ln p_t - \ln p_{t-1}$. Missing data were replaced with a zero log-return in order to get an homogeneous sample. Computation were performed in GAUSS version 6 on a desktop Pentium 3 personal computer, with clock frequency of 1 Ghz and a Ram of 256 Mbytes.

4.2 Conditional mean

In order to estimate $\mu_t = \mu w_t$ we fitted a multivariate regression model using the following explanatory variables: the lagged log-returns of the Standard & Poor's 500 Index, the lagged first difference of the interest rates on 3 Months Treasury Bills and on 10 Years Notes, the lagged log-difference of Oil Prices (Texas), a set of dummy variables for the day-of-the-week effect and the January effect.

For brevity OLS estimation outputs are not reported; the results provide little evidence of the January effect while the day-of-the-week effect appears to be more relevant, in particular for Monday. Furthermore, the lagged values first difference of the interest rate on 10 Years notes do not appear to be significant.¹ For simplicity we retained all regressors in the specification of the conditional mean, regardless of significance.

4.3 Estimation strategy for the conditional variance

We fitted three models: a simple CC with unrestricted R correlation matrix, a SEARCH model with RHOM S matrix and a SEARCH model with HOM S matrix. We use market capitalization as intra-class ordering criteria and we used specification (18) for spatial matrices in GARCH dynamics. In both models R_{ij} was estimated with the unrestricted specification (19).

All models were estimated by Quasi Maximum Likelihood (QML). In numerical optimization, we started iteration with a BFGS algorithm, for a maximum the first

¹These observations are based on the F-test for retained regressors on the full OLS unrestricted estimation.

20 iterations. We next switched to the Newton algorithm. The CC model was estimated using the two-step approach proposed by Bollerslev (1990).

The two SEARCH models have been estimated using an alternating maximization approach, see Caporin and Paruolo (2005). This consists in alternating maximization of the S matrix parameters and the ζ , E , A and R parameters. The first is estimated conditionally to the estimated patterns for the η_t conditional variances obtained by fixing ζ , E , A and R coefficients at the last estimated value. The second parameter group is estimated conditionally to last S estimates. The iterative algorithm switches between the two parameters subset until convergence is achieved.

From a computational point of view, the estimation of ζ , E , A and R coefficients is relatively fast given the small number of assets within each class \mathcal{C}_{ij} , 5 in the present application. On the other hand the estimation of S is time-consuming, with a complexity that increases with the total number of assets and the complexity of the S structure.

Each iteration CPU time for the numerical optimization for S parameters took around 20 minutes, with 10 iterations on average. This estimation step involves data on all 150 variables. For the EAR coefficients, with blocks of 5 variables each, one iteration required few seconds, for an average of 40 iterations.

Although non-negligible, computing time is within capabilities of current personal computers. We expect a maximum computing time of a few hours on current workstations, and a fraction of this times on personal computer of the next generation, thanks to the rapid development of computer technology.

The SEARCH estimates are described in the next 3 subsections. The CC specification is used only for model comparisons, reported in the next section.

4.4 S estimates

For brevity we report estimated coefficients in SEARCH specifications graphically whenever possible. All SEARCH coefficients were significant on the basis of standard t -test with standard errors based on an outer product of the gradient of the asymptotic variance covariance matrix at the maximum. This guarantees a consistent estimate of the standard errors also when using the alternating maximization algorithm.

Figure 1 reports the estimated coefficients for the two S specification, RHOM and HOM. The assets are ordered as in Table 1 proceeding column-wise, from left to right. Vertical lines separate different subsets of parameters: the first block contains the coefficients for the market factor, the second one the coefficient to the 3 levels of the size factor; the third one contains the coefficients for the 10 levels of the sector factor; finally the last 3 panels contain estimates of the effects of the levels of the interaction factor.

[Figure 1, Table 1]

Squares and dashes in Figure 1 represents the RHOM coefficient values which are restricted to be equal within each panel. Comparing the RHOM to HOM specifications, one observes strong evidence of heterogeneity in the coefficients. We hence next comment only the HOM specification.

The relation among assets belonging to L firms is more pronounced than the

relation between M and S assets, which is very low. Assets belonging to the BM, EN, FI and UT sectors present a sizable positive coefficient.

The interaction effects highlights relevant differences across size classes. In fact coefficients associated to interactions in L assets are generally negative compared to the large positive coefficients for M and S. The UT sector has very different coefficient patterns regardless of size.

4.5 EA estimates

We next consider the estimates of the conditional variance dynamics; Fig. 2 reports the ζ constants in the GARCH dynamics.

[Figure 2] [Tables 2, 3.]

The ζ coefficients are similar for the RHOM and HOM specifications. They vary by asset dimension: L firms have smaller ζ_{ij} coefficients than M and S. Furthermore, the SE, TE and UT sectors have higher coefficients compared to other sectors. Finally, CG-M and CNG-M assets have higher conditional variance constants compared to the corresponding S assets.

These results are in line with expectations: *ceteris paribus* L firms should evidence a lower stock price volatility compared to M and S firms belonging to the same sector. Across sectors, SE and TE show high volatility in the period.

We next calculated roots of $|I - Ez - Az|$ implied by the numerical estimates of the GARCH matrices E and A for the HOM specification. All the calculated roots were real, and ranged from 1.02 to 6.36. Table 2 reports the frequency distribution of the inverted roots $1/z$. Some of these roots are close to 1, as found in many empirical studies of daily financial data. Note that the number of inverted roots close to 1 is limited.

Table 3 reports estimates of the E and A coefficients, both for the SEARCH RHOM and HOM specifications. The two specifications present only minor differences. The spatial coefficients included in EA are generally small, in particular for the A_{ij} matrix, although significant.

4.6 R estimates

We next report estimates for the R matrix, which has a block diagonal Toeplitz structure, with 5×5 blocks. Figure 3 reports the estimated R matrices coefficients for the RHOM and HOM specifications.

The fact that HOM has a more flexible specification for S has a relevant impact on the R coefficient estimates, apart from most of the L assets (where there are relevant changes for the SE, TE and UT sectors). The HOM specification provides mostly negative coefficients for M and S firms. As for the S matrix, the UT sector behaves differently and has positive coefficients.

[Figures 3 and 4]

Both R and S explain the correlations between assets. S accounts for between-classes effects of the market, size, sector while R explains within-group effects. It is thus of little surprise that a more flexible specification for S results in different correlations in R .

Figure 4 reports the $\rho_{ij,q}$ coefficients for the HOM specification, where each group of symbols refers to a different q . It can be seen that for fixed i, j , the $\rho_{ij,q}$ are approximately equal. This thus seems to suggest a spatial correlation structure within each class \mathcal{C}_{ij} of the type where all the assets belonging to the class are first order neighbors, see Subsection 3.7.

This observation can lead to further parameter restrictions, obtained by imposing $\rho_{ij,q} = \rho_{ij}$. This would reduce the number of parameters in R from 120 to 30. The HOM specification could be further refined. Restriction on the parameters can be imposed, grouping assets with similar coefficient behavior. For instance SE and TE sectors have very similar coefficient patterns in S and R , as well as the BM and EN sectors. Further reductions in the R matrix structure may be considered. However, already with the restriction $\rho_{ij,q} = \rho_{ij}$, the number of parameters in R is very small.²

For simplicity none of these restrictions was imposed on the model, prior to the model evaluation reported in the next section.

5 Model evaluation

In this section we present model evaluation both for the SEARCH HOM specification and for the CC model. Results for the OG specification were similar to the one of CC and they are not reported for brevity.³ Subsection 5.1 reports residual analysis. Subsection 5.2 evaluates relative performances based on indicators derived from model-based portfolio management practices. Model comparison is performed both in sample and out of sample. Finally Subsection 5.3 reports the number of Value-at-risk exceptions obtained for the model-based portfolios and on a volatility forecast indicator, related to the suggestions in Engle and Colacito (2003).

5.1 Residual analysis

Models were analyzed in order to verify the presence of autocorrelations in the squared standardized residuals. In a first step we applied a univariate Ljung-Box test. The test outcomes are summarized in Table 4; they show that in-sample the CC model produces slightly better results, resulting in a lower number of rejections of the null hypothesis. On the contrary both models provide comparable results out-of-sample with a small preference for the SEARCH model. Overall univariate evidence appears in line with correct specification of both models.

[Table 4]

We computed also a multivariate Ljung-Box test, see Hosking (1980), and the omnibus test for normality, see Doornik and Hansen (1994). The two tests reject the null hypothesis of absence of autocorrelation and normality, respectively. This

²Other restrictions and extension can be considered on the GARCH dynamics. For instance one can consider the E matrix to be diagonal ($m_{E_{ij,l}} = 0$) and leave the blocks A_{ij} in the A matrix completely unrestricted. In this case in particular, on the GARCH step of the alternating algorithm, one can estimate conditional variances $h_{ij,l}$ corresponding to each asset separately, exploiting the diagonality of E .

³The results are available from the authors on request.

is a common finding in applied ARCH modeling. The joint univariate and multivariate results are interpreted here as evidence of departures from normality of the conditional distribution of returns. The present estimates are thus to be interpreted as QML estimates.

5.2 Implied portfolio management

This subsection reports model comparison for the two SEARCH models and the CC, both in and out of sample. In order to provide a meaningful forecast performance indicator, we consider an asset allocation framework, as in Engle and Colacito (2003). Portfolio managers usually prefer investment strategies that are least volatile over time, and highest performance indices. We hence tried to document these features for portfolios formed on the basis of the SEARCH and CC 1-step ahead predictions.

We considered two portfolio optimization strategies: the mean-variance optimal portfolio with or without short sales constraints. The expected desired mean asset return was set equal to $n^{-1}i'_n\mathbb{E}_t(y_{t+1})$, i.e. the average mean forecast in the cross-section of assets. The specification of the conditional expectation $\mathbb{E}_t(y_{t+1})$ is the same for the CC and SEARCH models, so that differences in portfolio weights reflect only differences in prediction of the conditional variance-covariance matrix $\mathbb{V}_t(y_{t+1})$.

Portfolio weights were re-calculated every day using the model prediction for $\mathbb{E}_t(y_{t+1})$ and $\mathbb{V}_t(y_{t+1})$ assuming no transaction costs. There are 4 combinations of forecasts (SEARCH and CC) by portfolio allocation scheme (with short sales, WSS, no short sales, NSS).

Figures 5, 6 report weights by sector in the period January 1999 to June 2001 for all model-allocation combinations. Figures 7 graphs weights by size. The same information for the case of no short sales is collected in Figures 8, 9 for sector weights, and in Figure 10 for size. Table 5 and 6 report summary statistics for the in-sample period (the last 500 observations which cover the years 1998 and 1999) and out-of-sample (January 2000 to June 2001).

[Tables 5 and 6, Figures 5, 6, 7, 8, 9, 10]

We first analyze volatility of the portfolio weights. Weight variability, as represented by the in- and out-of sample standard deviation, is greater for CC portfolio weights than for the SEARCH portfolio weights, in the case of no short sales. The opposite is the case for the case with short sales. In general, however, the CC portfolio weights show more persistent departures from the average than the SEARCH portfolio weights. In this sense SEARCH appears to deliver more stable weights.

The importance of different sectors is model-allocation specific. For most model-allocation pairs, UT is the most important sector, followed by CNG, BM and CG, with these four sectors accounting for about 70% of all portfolios. SEARCH assigns larger weights to the NCG than CC; furthermore, SEARCH seems to take more advantage from short-sales opportunities than CC.

Comparing in-sample and out-of-sample portfolio compositions, one observes relevant changes in CC weights on average: the weight of BM decreases from 15.4% to 10.2%, the one of UT from 40.4% to 36%, while the one of CG increases from 8.3% to 11.3%, in case with short sales. Comparably, SEARCH weights appear more

stable, with less variation in average weights comparing in-sample and out-of-sample behavior.

Considering portfolio composition by size, one sees that L firms have larger weights in the CC portfolio than in SEARCH's, where more weight is assigned to M firms. Overall it appears that SEARCH based strategies show a less volatile performance than CC based strategies.

[Figures 11, 12, Tables 7, 8]

We next turn attention to overall performance of portfolios. The compounded returns on portfolios with active day-by-day re-balancing were calculated, see figure 11. Portfolio realized returns have common patterns in the four model-allocation schemes. SEARCH portfolios provide higher compounded returns both in-sample that out-of-sample (see Figure 11 and Table 7), except in one case. Hence SEARCH based strategies appear to deliver superior performances. SEARCH portfolio realized variances are also higher, see Table 8 and Figure 12, in line with predictions of mean-variance analysis.

5.3 Volatility performance indicators

In this subsection we consider volatility performance indicators for model comparison. We consider value at risk exceptions for the portfolios constructed in the previous subsection and the ratios of realized versus predicted portfolio volatility, in line with the suggestion in Engle and Colacito (2003).

The number of Value-at-risk exceptions for the various model-allocation pairs, in- and out-of-sample, is reported in Table 9. The table shows a better performance of the SEARCH model.

We next consider the approach of Engle and Colacito (2005). They consider the ratio EC_{t+1} between the 'realized portfolio variance', represented by $(\hat{w}_t' \hat{\varepsilon}_{t+1})^2$, and 'predicted portfolio variance', represented by $\hat{w}_t' \hat{\Sigma}_t \hat{w}_t$, where $\hat{\varepsilon}_t := y_t - \hat{\mathbb{E}}_{t-1}(y_t)$. Note that for both CC and SEARCH $\hat{\mathbb{E}}_{t-1}(y_t)$ and $\hat{\varepsilon}_t$ are the same.

As portfolio weights \hat{w}_t Engle and Colacito used the ones corresponding to WSS in the present notation. Results are reported in Table 10, which also collects results for NSS weights. Both WSS and NSS variance weights, depend on (the same) prediction for the conditional mean $\hat{\mathbb{E}}_{t-1}(y_t)$. In order to eliminate dependence on $\hat{\mathbb{E}}_{t-1}(y_t)$ one can use instead the global minimum variance weights $\hat{w}_t := \hat{\Sigma}_t^{-1} \iota (\iota' \hat{\Sigma}_t^{-1} \iota)^{-1}$, which do not depend on $\hat{\mathbb{E}}_{t-1}(y_t)$. The resulting time series of $\hat{w}_t' \hat{\Sigma}_t \hat{w}_t$ for CC and SEARCH weights are pictured in Fig. 13.

[Tables 9, 10, Fig. 13.]

For a correctly specified volatility model, one expects EC_{t+1} to be approximately centered around 1. When comparing alternative volatility models, one should prefer the one with ratio closer to one. Table 10 reports summary statistics of this ratio for the different models, both in-sample and out-of-sample. The SEARCH models provide average EC_{t+1} ratios closer to one out-of-sample, with comparable in-sample performances of SEARCH and CC. Also on this count, SEARCH models appears to outperform competitors.

6 Conclusions

In this paper we have applied a spatial multivariate ARCH specification, called SEARCH, which employs spatially restricted parameter matrices; this class of models has number of parameters that is linear in the number of assets.

The application to daily returns on 150 stocks from the NYSE for the period January 1994 to June 2001 showed that this model is estimable on a personal computer also for a large cross section. Parameter estimates are interpretable and lend themselves to several specification searches strategies.

These models were evaluated against a standard CC model on several counts. The in- and out-of-sample performance of portfolio allocation strategies based on the SEARCH model outperform the ones based on CC. The corresponding SEARCH portfolio weights appear also to be more stable over time than the CC counterparts.

The percentage of risk exceptions and the ratios of realized and predicted volatility suggested in Engle and Colacito (2005) also confirms a better performance of SEARCH models compared to CC. Although empirical findings cannot be extended to other data-sets, the present empirical application suggests that spatial models may provide a sensible approach to parsimonious volatility modeling.

References

- Anselin L 1988. *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers: Dordrecht, the Netherlands.
- Alexander C 2001. Orthogonal GARCH. In *Mastering Risk*. Financial Times–Prentice Hall: London.
- Amisano G, Giannini C, 1997. *Topics in structural VAR econometrics*. 2nd revised and enlarged edition. Springer Verlag: Berlin.
- Baltagi B H, Song S H, Jung B C, Koh W, 2003. Testing for Serial Correlation, Spatial Autocorrelation and Random Effects Using Panel Data. Working paper. Available at <http://repec.org/esAUSM04/up.3749.1080795402.pdf>
- Bauwens L, Laurent S and Rombouts JKV, 2004. Multivariate GARCH models: a survey. CORE discussion paper 2003/31. Available at <http://www.core.ucl.ac.be/services/COREdp03.html>
- Bollerslev T, 1990. Modeling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH approach. *Review of Economics and Statistics* **72**: 498–505.
- Caporin M, Paruolo P, 2005. Spatial effects in multivariate ARCH. Quaderno della Facoltà di Economia 2005/1, University of Insubria. Available at <http://eco.uninsubria.it/dipeco/quaderni/y2005.htm>
- Case AC, 1991. Spatial patterns in household demand. *Econometrica* **59**: 953–965.
- Christoffersen PF, 1998. Evaluating interval forecasts, *International Economic Review* **39**: 841–862.
- Conley TG, Topa G, 2002. Socio-economic distance and spatial patterns in unemployment. *Journal of Applied Econometrics*: **17**: 303–327. DOI: 10.1002/jae.670.
- Cressie NAC, 1993. *Statistics for spatial data*, revised edition. John Wiley and Sons: New York.

- Doornik JA, Hansen H, 1994. An Omnibus Test for Univariate and Multivariate Normality. Working paper. Available at <http://www.doornik.com/research/normal2.pdf>.
- Engle R, Colacito R, 2003. Testing and Valuing Dynamic Correlations for Asset Allocation. Working paper. Available at <http://www.econ.upenn.edu/cgi-bin/mecon/bin/serve.cgi/TestingandValuingMarch291>
- Engle R, Kroner K. 1995. Multivariate simultaneous generalized ARCH. *Econometric Theory* **11**: 122–150.
- Giacomini R., C. W.J. Granger 2004. Aggregation of space-time processes. *Journal of Econometrics* **118**: 7–26. DOI:10.1016/S0304-4076(03)00132-5
- Hosking JRM, 1980. The multivariate portmanteau statistic. *Journal of the American Statistical Association* **75**: 602-608.
- Ling S, McAleer M, 2003. Asymptotic theory for a new vector ARMA-GARCH model. *Econometric Theory* **19**: 280-310. DOI: 10.1017/S0266466603192092
- Ling S, Li WK, 1997. Diagnostic checking on nonlinear multivariate time series with multivariate ARCH errors, *Journal of Time Series Analysis* **18**: 447-464.
- Tse YK, 2002. Residual based diagnostics for conditional heteroscedasticity models, *Econometrics Journal* **5**: 358-373.
- Tse YK, AKC Tsui, 1999. A note on diagnosing multivariate conditional heteroscedasticity models. *Journal of Time Series Analysis* **20**: 679-691.

A Appendix: List of assets by sector and size

BM-L: E.I. Du Pont & Co., Dow Chemical Co., Kimberly-Clark Corp., Alcoa Inc., Newmont Mining Corp.

BM-M: The Valspar Corp., Sonoco Products Co., The Scotts Co., Bowater Inc., Lubrizol Corp.

BM-S: HB Fuller Co., Rogers Corp., Universal Forest Prod. Inc., Caraustar Industries Inc., Chesapeake Corp

CG-L: Honeywell Int. Inc., Illinois Tool Works Inc., Lockheed Martin Corp., Northrop Grumman Corp., Deere and Co.

CG-M: Tell Brothers Inc., SPX Corp., Alliant Techsystems Inc., Hovnanian Enterprises Inc., Jacobs Engineering Group Inc.

CG-S: Fletwood Enterprises Inc., Thomas Industries Inc., Stewart & Stevenson Serv. Inc., Apogee Enterprises Inc., Pope & Talbot Inc.

CDG-L: Ford Motor Co., General Motors Corp., NIKE Inc., Harley-Davidson Inc., Masco Corp.

CDG-M: Mohawk Industries Inc., Gentex Corp., Borg Warner Inc., HNI Corp., Timberland Co.

CDG-S: K2 Inc., Applied Industrial Techn. Inc., Gencorp. Inc., TBC Corp., Interface Inc.

CNG-L: Altria Group, PepsiCo, Colgate-Palmolive Co., Avon Products, Kellogg Company

CNG-M: Constellation Brands, Hormel Foods Corporation, Smithfield Foods, PepsiAmericas, Church & Dwight Co.

CNG-S: Sanderson Farms, WD-40 Company, Lance, J&J Snack Foods Corp., Standard Register Com

EN-L: Exxon Mobil Corp., ChevronTexaco Corp., Devon Energy Corp., Anadarko Petroleum Corp., Apache Corp.

EN-M: Lyondell Chemical Company, Patterson-UTI Energy, Pogo Producing Company, Pride International, Western Gas Resources Inc.

EN-S: Offshore Logistics, Veritas DGC Inc., TETRA Technologies, Atwood Oceanics, Frontier Oil Corp.

FI-L: American International Gr., Fannie Mae, Countrywide Financial Corp., AFLAC Incorporated, Aetna Inc.

FI-M: Old Republic International, AmeriCredit Corp., Bank of Hawaii Corp., FirstMerit Corporation, IndyMac Bancorp

FI-S: Irwin Financial Corp., First Republic Bank, Cash America International, Presidential Life Corp., World Acceptance Corp.

HE-L: Johnson & Johnson, Abbott Laboratories, Medtronic, Wyeth, HCA Inc.

HE-M: Varian Medical Systems, Barr Pharmaceuticals, DENTSPLY International Inc., Lincare Holdings Inc., Universal Health Services

HE-S: Immucor, Cambrex Corporation, Enzo Biochem, RehabCare Group, Bradley Pharmaceuticals

SE-L: Time Warner Inc., Lowe's Companies, Inc., Starbucks Corporation, Best Buy Co., Inc., Staples, Inc.

SE-M: Whole Foods Market, Inc., Urban Outfitters, Inc., Claire's Stores, Inc., O'Reilly Automotive, Inc., Applebee's Int'l, Inc.

SE-S: Lone Star Steakhouse, Steak n Shake Company, Bowne & Co., Inc., J. Jill Group, Inc., Nash Finch Company

TE-L: Dell Inc., Hewlett-Packard Company, First Data Corporation, Symantec Corporation, Electronic Arts Inc.

TE-M: McAfee, Inc., Synopsys, Inc., Tech Data Corporation, Ametek, Inc., Cree, Inc.

TE-S: Baldor Electric Company, A.O. Smith Corporation, Aeroflex Incorporated, Cubic Corporation, C-COR.net Corporation

UT-L: Exelon Corporation, Southern Company, Dominion Resources, Duke Energy Corporation, Entergy Corporation

UT-M: SCANA Corporation, Alliant Energy Corporation, DPL Inc., National Fuel Gas Co., AGL Resources Inc.

UT-S: El Paso Electric Company, CH Energy Group, UIL Holdings Corporation, Laclede Group, American States Water Com.

B Graphs and tables

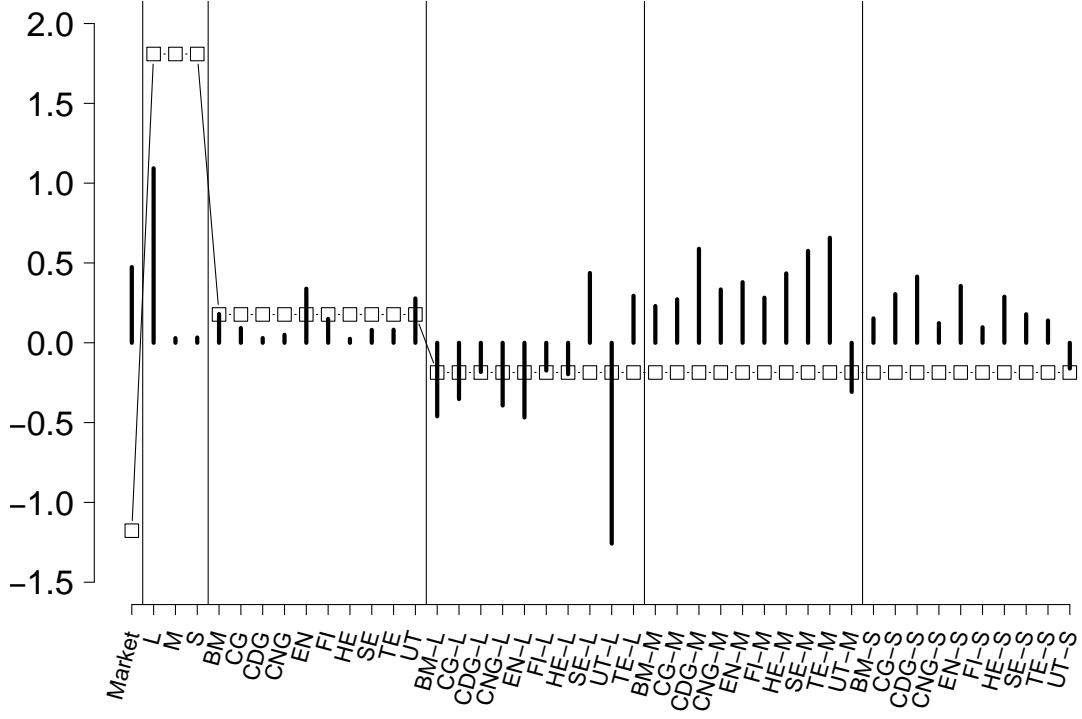


Figure 1: Coefficients of the S matrix. Bars: HOM specification, squares and solid line: RHOM specification.

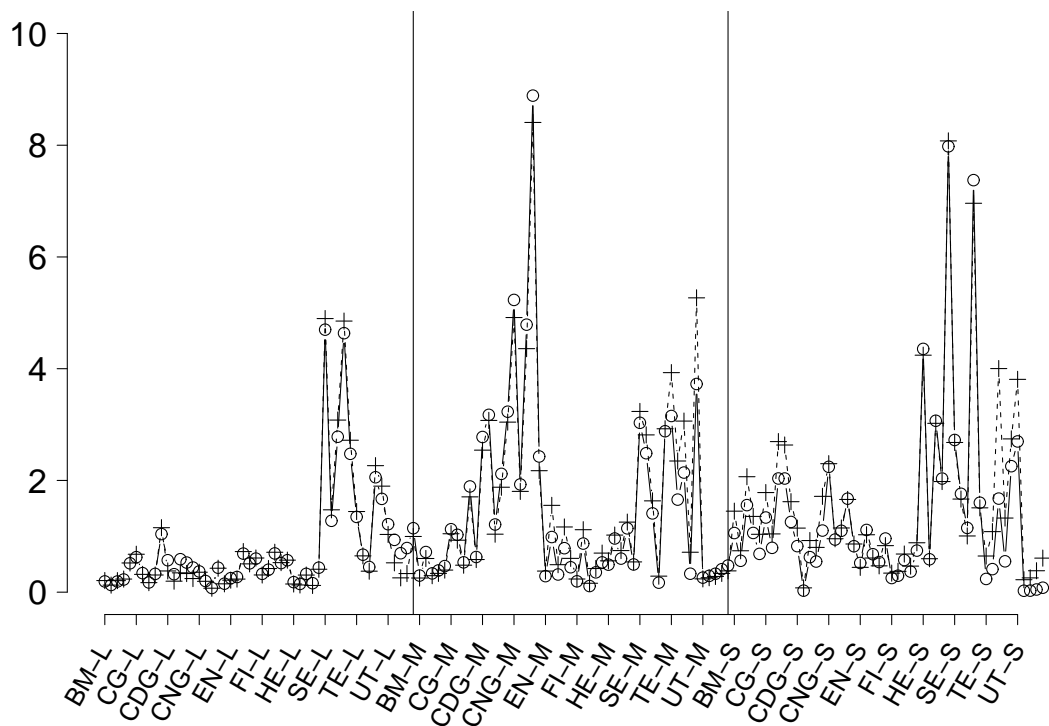


Figure 2: Coefficients ζ , intercepts in GARCH equations. Crosses and dashes: RHOM specification; circles and solid line: HOM specification.

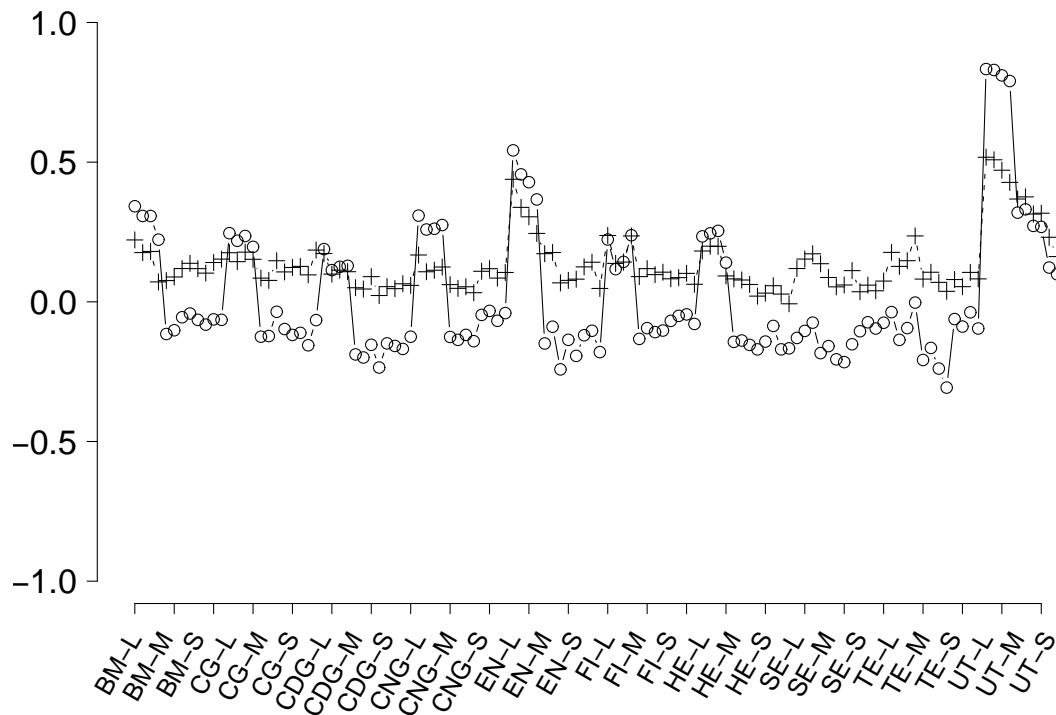


Figure 3: Coefficients in R . Crosses and dashes: RHOM specification; circles and solid line: HOM specification.

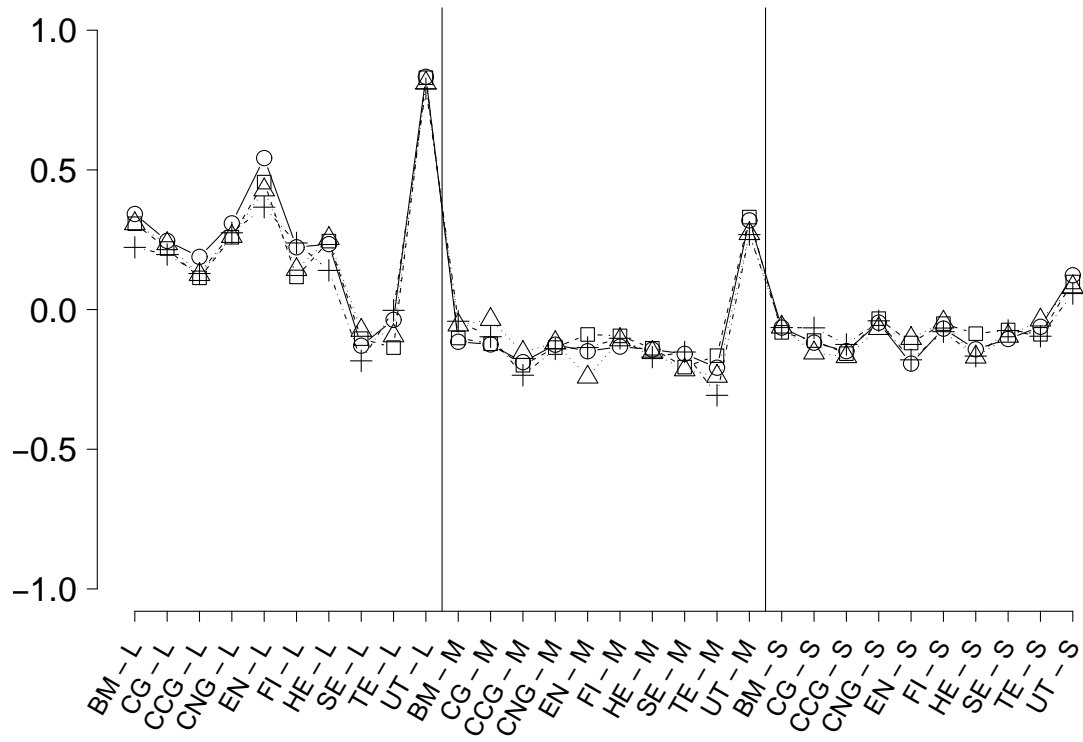


Figure 4: Coefficients in R for the HOM specification. Circles and solid line: $\rho_{ij,1}$; squares and dashes: $\rho_{ij,2}$; triangle and dotted line: $\rho_{ij,3}$; pluses and dotdashed line: $\rho_{ij,4}$.

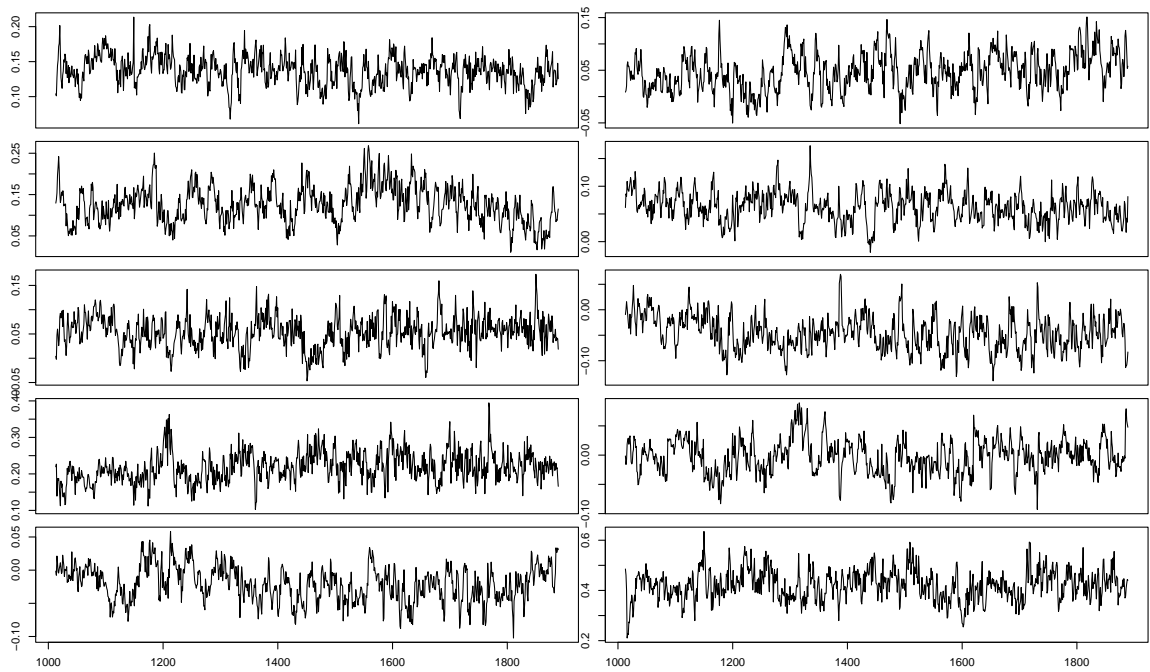


Figure 5: SEARCH-WSS: portfolio weights with short sales, by sector. Sectors are ordered column-wise, top to bottom. Estimation sample ends at observation 1513.

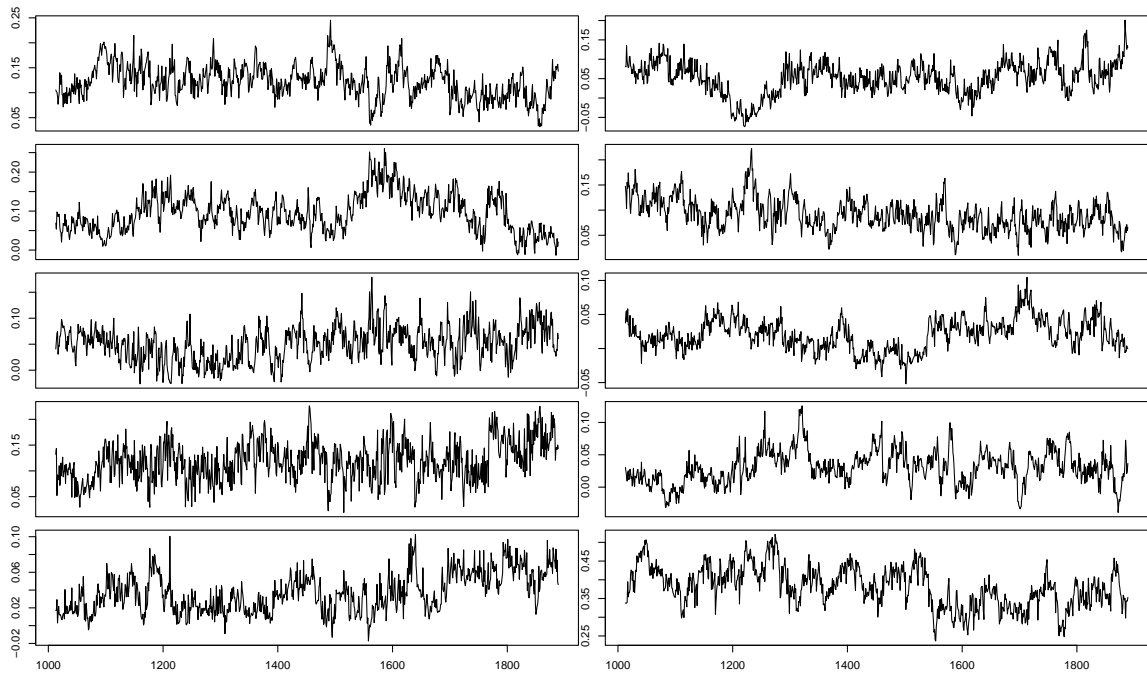


Figure 6: CC-WSS: portfolio weights with short sales, by sector. Sectors are ordered column-wise, top to bottom. Estimation sample ends at observation 1513.

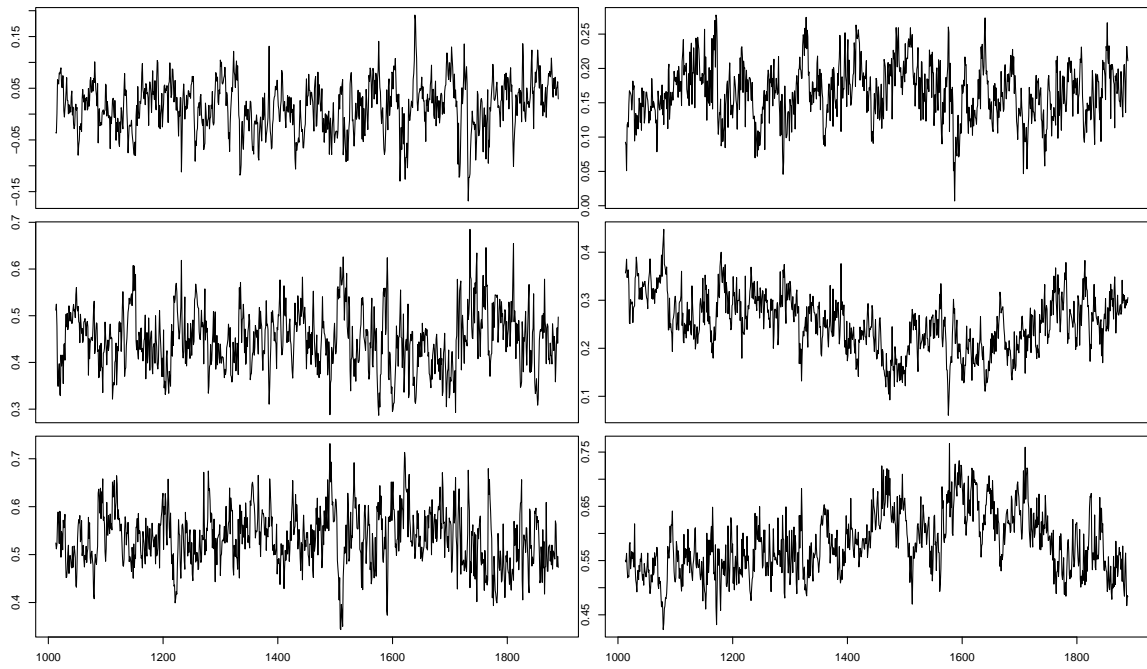


Figure 7: SEARCH-WSS and CC-WSS: portfolio weights with short sales, by size. Left column SEARCH, right column CC; rows (top to bottom): L S M. Estimation sample ends at observation 1513.

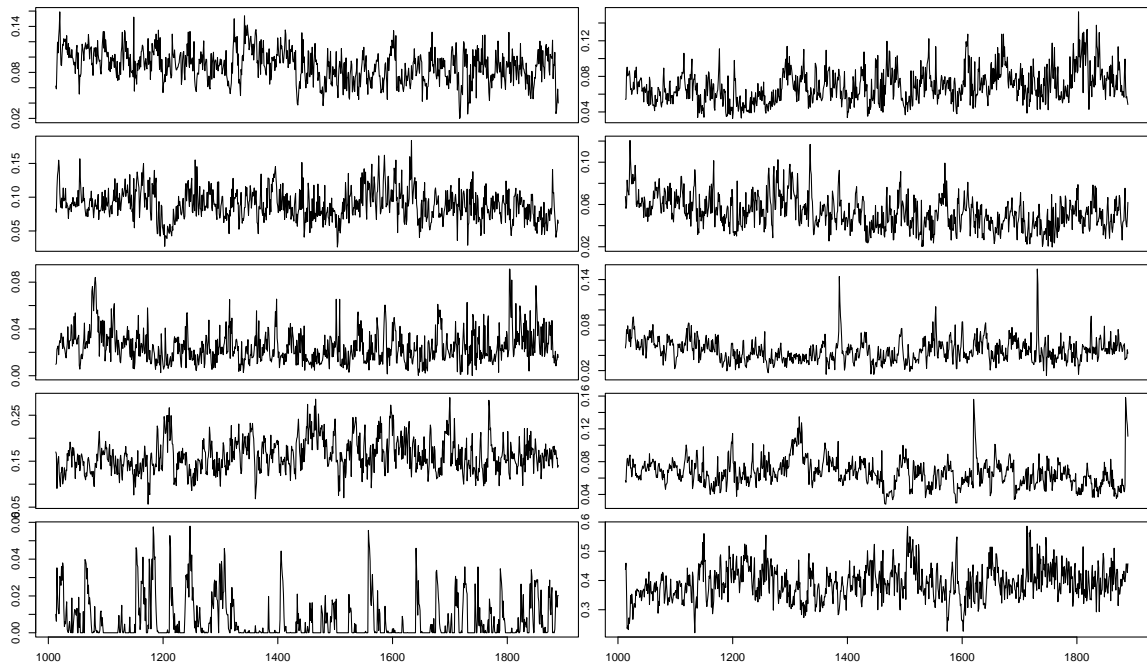


Figure 8: SEARCH-NSS: portfolio weights with no short sales, by sector. Sectors are ordered column-wise, top to bottom. Estimation sample ends at observation 1513.

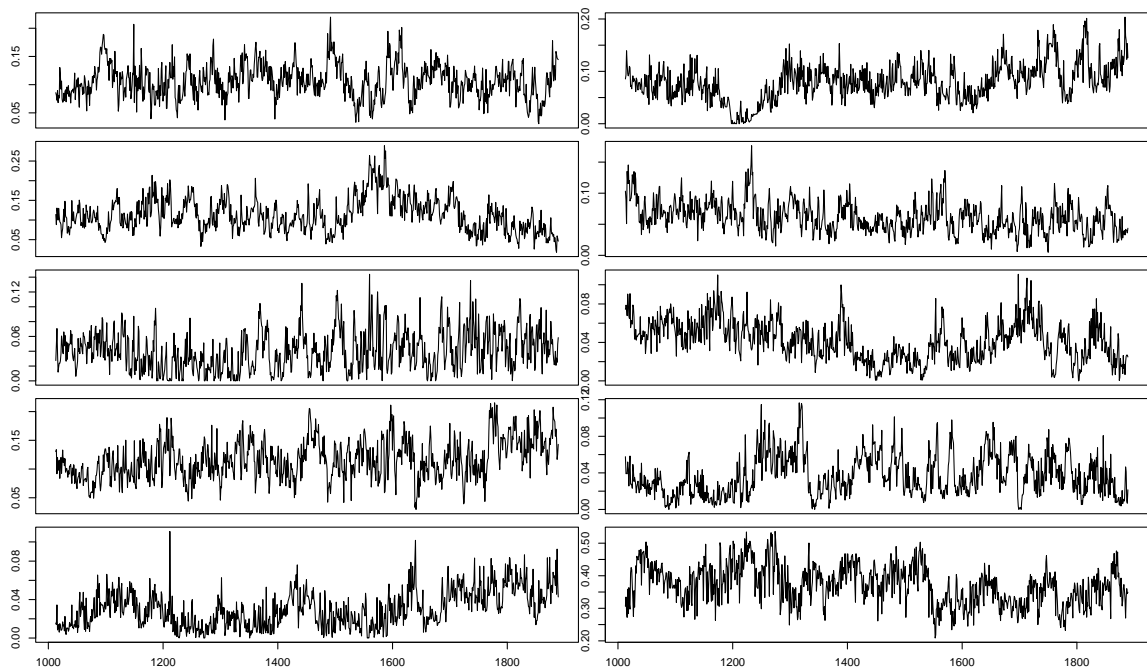


Figure 9: CC-NSS: portfolio weights with no sales, by sector. Sectors are ordered column-wise, top to bottom. Estimation sample ends at observation 1513.

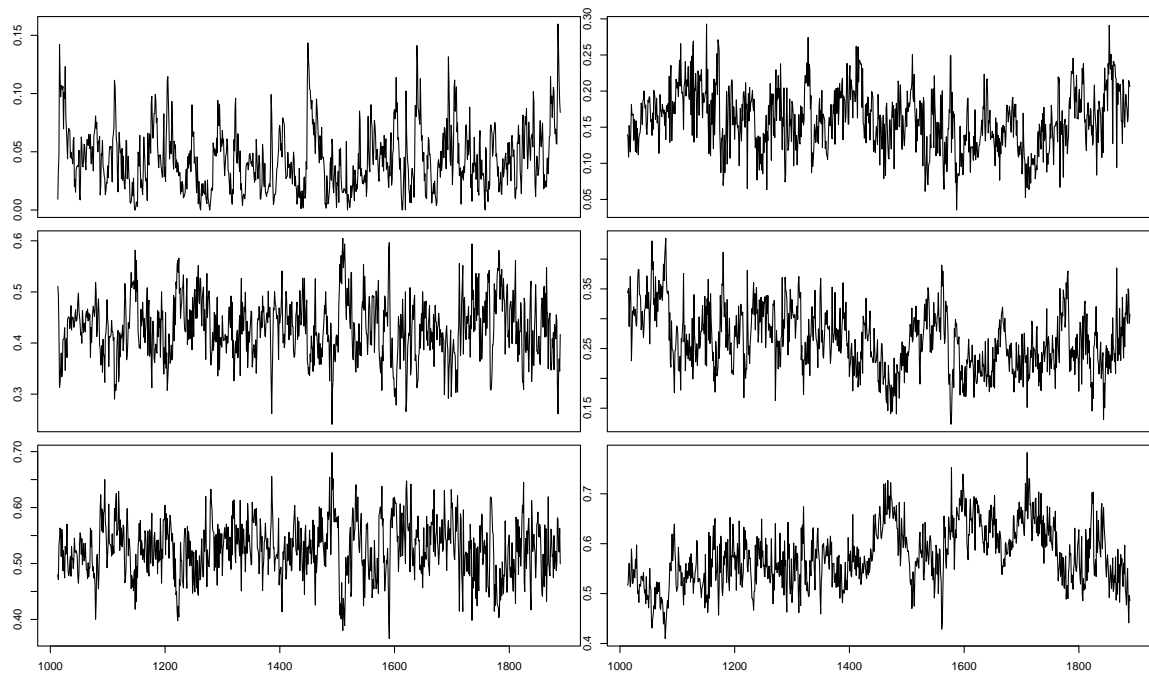


Figure 10: SEARCH-NSS and CC-NSS: portfolio with no short sales, by size. Left column SEARCH, right column CC; rows (top to bottom): L S M. Estimation sample ends at observation 1513.

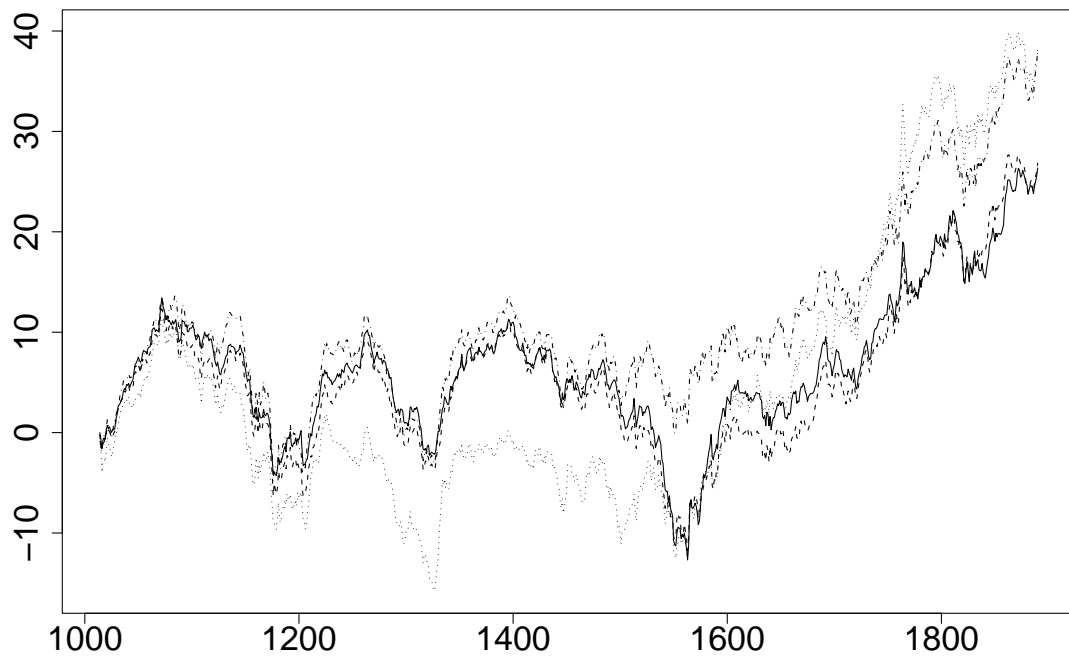


Figure 11: Compounded returns on CC and SEARCH portfolios. CC-WSS: dashes, CC-NSS: solid, SEARCH-WSS: dot-dashes, SEARCH-NSS: dots. Estimation sample ends at observation 1513.

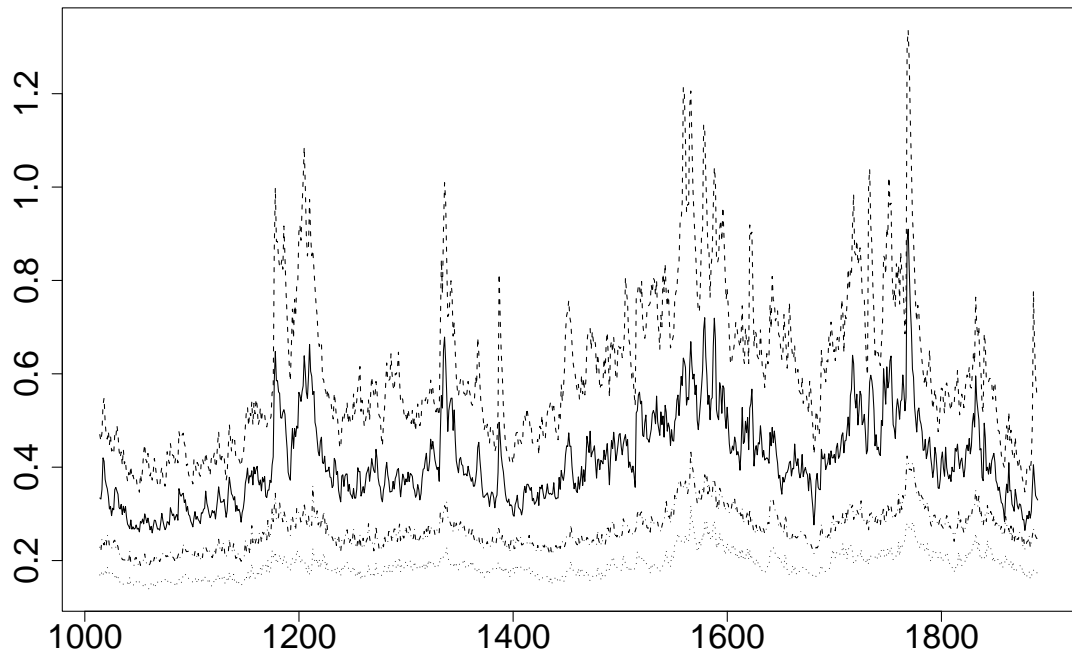


Figure 12: Conditional variances on SEARCH and CC portfolios. SEARCH-NSS: dashes, SEARCH-WSS: solid, CC-NSS: dot-dashes, CC-WSS: dots. Estimation sample ends at observation 1513.

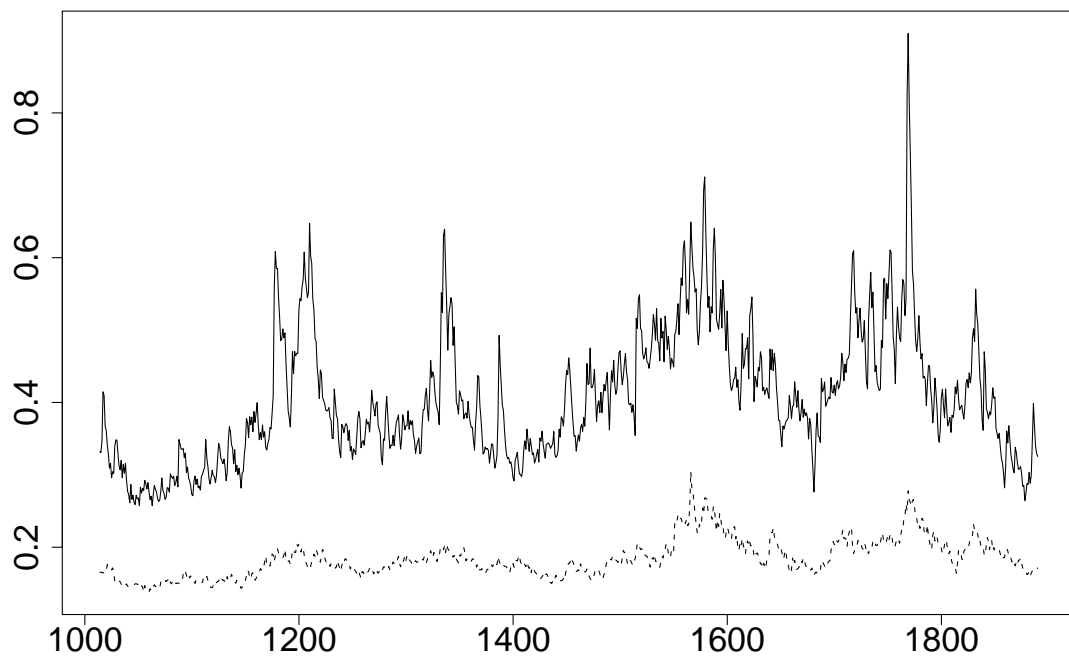


Figure 13: Conditional variances for SEARCH and CC minimum variance portfolios. SEARCH: solid, CC: dashes. Estimation sample ends at observation 1513.

Capitalization level \rightarrow Industrial sector \downarrow	level 1	level 2	...	level ℓ
sector 1	\mathcal{C}_{11}	\mathcal{C}_{12}	...	$\mathcal{C}_{1\ell}$
...
sector k	\mathcal{C}_{k1}	\mathcal{C}_{k2}	...	$\mathcal{C}_{k\ell}$

Table 1: Two-way classification based on industrial sector and capitalization size.

classes	$(0, 0.80]$	$(0.80, 0.85]$	$(0.85, 0.90]$	$(0.90, 0.95]$	$(0.95, 0.98]$	total
number	68	18	26	32	6	150

Table 2: Inverted roots $1/z$ of the GARCH equations, HOM specification.

i	j	RHOM				HOM			
		$\beta_{ij,10}$	$\beta_{ij,11}$	$\alpha_{ij,10}$	$\alpha_{ij,11}$	$\beta_{ij,10}$	$\beta_{ij,11}$	$\alpha_{ij,10}$	$\alpha_{ij,11}$
BM - L		0.83	0.00	0.09	0.01	0.85	0.00	0.08	0.01
	- M	0.83	0.00	0.09	0.00	0.81	0.00	0.10	0.00
	- S	0.54	0.04	0.11	0.00	0.61	0.03	0.12	0.00
CG - L		0.38	0.18	0.12	0.03	0.39	0.16	0.12	0.04
	- M	0.65	0.00	0.14	0.00	0.60	0.00	0.16	0.00
	- S	0.42	0.00	0.23	0.03	0.47	0.00	0.22	0.03
CCG - L		0.86	0.00	0.06	0.00	0.77	0.01	0.10	0.00
	- M	0.58	0.00	0.09	0.01	0.56	0.00	0.08	0.01
	- S	0.43	0.00	0.29	0.12	0.49	0.00	0.29	0.13
CNG - L		0.77	0.03	0.11	0.00	0.78	0.04	0.10	0.00
	- M	0.21	0.00	0.10	0.00	0.13	0.00	0.10	0.00
	- S	0.60	0.00	0.12	0.01	0.57	0.00	0.14	0.00
EN - L		0.76	0.00	0.08	0.02	0.78	0.00	0.07	0.02
	- M	0.80	0.00	0.10	0.00	0.86	0.00	0.08	0.00
	- S	0.88	0.00	0.06	0.00	0.86	0.00	0.07	0.00
FI - L		0.70	0.00	0.13	0.02	0.70	0.00	0.12	0.02
	- M	0.73	0.01	0.14	0.00	0.75	0.01	0.15	0.00
	- S	0.83	0.00	0.09	0.00	0.84	0.00	0.09	0.00
HE - L		0.83	0.01	0.06	0.02	0.81	0.01	0.07	0.02
	- M	0.68	0.00	0.16	0.00	0.69	0.01	0.16	0.00
	- S	0.57	0.01	0.18	0.00	0.57	0.01	0.17	0.00
SE - L		0.52	0.01	0.13	0.01	0.50	0.04	0.11	0.00
	- M	0.47	0.07	0.16	0.02	0.47	0.10	0.13	0.01
	- S	0.56	0.01	0.19	0.00	0.52	0.01	0.22	0.00
TE - L		0.75	0.00	0.11	0.01	0.77	0.00	0.09	0.01
	- M	0.70	0.00	0.10	0.00	0.79	0.01	0.04	0.00
	- S	0.59	0.00	0.12	0.00	0.80	0.00	0.09	0.00
UT - L		0.32	0.13	0.10	0.01	0.54	0.03	0.08	0.01
	- M	0.70	0.00	0.08	0.02	0.57	0.00	0.11	0.03
	- S	0.74	0.00	0.08	0.00	0.95	0.00	0.02	0.00

Table 3: Estimates of the E_{ij} and A_{ij} matrices in GARCH dynamics, for the HOM and RHOM SEARCH specifications.

L:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	C	C																		
I:	16	8	7	6	6	7	6	6	5	6	5	5	5	5	7	6	6	7	8	8
O:	16	15	12	14	12	12	12	10	8	9	9	8	8	9	9	8	8	8	8	7
	S	E	A	R	C	H														
I:	17	14	12	13	14	14	13	14	12	12	15	14	15	16	14	18	18	21	19	18
O:	13	11	8	10	11	8	8	8	9	9	9	9	8	9	9	7	7	7	7	7

Table 4: Number of rejections of the null hypothesis in Ljung-Box test at the 5% level out of 150 assets. L: lag, 1 to 20. I: in-sample period, 500 observations from January 1998 to December 1999. O: Out of sample period, 377 observations from January 2000 to June 2001.

	In sample				Out of sample			
SEARCH	100 <i>m</i>	100 <i>s</i>	100min	100max	100 <i>m</i>	100 <i>s</i>	100min	100max
BM	18.4	2.6	10.0	24.9	17.4	2.2	11.4	23.2
CG	13.6	3.8	5.1	24.6	14.1	4.9	3.2	28.5
CCG	2.3	2.9	-6.4	8.9	2.5	2.4	-6.1	9.2
CNG	22.0	4.0	11.8	33.2	23.7	3.8	13.1	35.4
EN	-1.4	2.6	-9.1	5.2	-2.7	2.5	-10.2	4.4
FI	4.8	3.3	-2.9	13.9	6.3	2.9	-1.2	14.3
HE	8.1	2.6	-0.5	14.8	7.6	2.4	2.0	16.2
SE	-4.3	3.2	-12.9	4.4	-5.7	3.3	-15.4	2.6
TE	-0.4	3.0	-8.9	8.3	-0.4	2.6	-6.8	8.1
UT	36.9	5.5	23.3	59.5	37.2	5.2	24.0	51.6
L	0.7	4.8	-14.9	13.6	3.2	5.1	-14.2	16.9
M	49.1	5.3	34.6	65.8	49.0	7.2	30.1	71.0
S	50.2	4.7	31.2	65.1	47.8	5.8	31.5	63.8
CC								
BM	15.4	3.0	7.1	22.9	10.2	3.5	2.6	19.7
CG	8.3	4.0	-0.1	20.2	11.3	5.9	-2.0	29.6
CCG	2.1	3.0	-6.1	10.3	3.9	2.9	-3.5	12.7
CNG	13.3	3.5	4.2	21.1	15.7	4.6	3.2	26.3
EN	2.3	1.9	-1.2	9.8	3.7	2.1	-1.4	9.4
FI	2.9	3.5	-8.1	12.2	3.8	3.0	-3.3	13.2
HE	11.6	3.5	4.4	22.8	10.7	2.7	2.8	18.7
SE	1.2	2.1	-4.6	7.7	2.7	2.3	-3.0	10.3
TE	2.6	2.8	-4.0	12.3	2.0	2.3	-4.0	9.1
UT	40.4	5.2	25.2	62.3	36.0	5.2	22.0	51.1
L	15.0	4.3	2.6	25.8	13.6	3.7	4.9	23.8
M	29.2	6.4	9.9	48.2	27.8	6.1	10.6	45.4
S	55.9	5.2	42.6	73.5	58.6	6.4	44.8	75.1

Table 5: Summary statistics for the portfolio weights implied by SEARCH and CC allowing for short sales. In sample: January 1998 - December 1999. Out of sample: January 2000 - June 2001. *m*: sample average. *s*: sample standard deviation. min,max: sample minimum and maximum.

	In sample				Out of sample			
SEARCH	100 <i>m</i>	100 <i>s</i>	100min	100max	100 <i>m</i>	100 <i>s</i>	100min	100max
BM	9.5	2.1	3.7	15.9	8.2	2.0	2.0	13.4
CG	8.9	2.3	2.6	15.7	8.9	2.4	2.9	18.4
CDG	2.5	1.4	0.3	8.4	2.5	1.5	0.0	9.1
CNG	16.0	3.7	5.6	28.5	16.6	3.6	7.0	28.8
EN	0.8	1.2	0.0	5.8	0.6	1.0	0.0	5.6
FI	6.4	1.7	3.3	11.9	7.6	2.0	3.8	15.3
HE	5.7	1.6	2.4	12.1	4.7	1.3	2.0	9.9
SE	4.5	1.5	1.5	14.4	4.8	1.5	1.3	15.4
TE	7.0	1.6	2.8	13.5	6.2	1.8	3.0	15.8
UT	38.8	5.8	22.1	58.4	39.8	5.8	22.6	58.6
L	4.2	2.7	0.0	14.3	4.7	2.7	0.0	16.0
M	43.1	5.5	24.1	60.5	42.8	6.4	26.2	59.6
S	52.6	4.7	38.0	69.8	52.5	5.4	36.6	64.8
CC								
BM	11.0	2.9	3.7	22.0	10.2	3.1	3.1	20.2
CG	10.8	3.4	3.3	21.4	11.9	5.1	1.8	28.9
CDG	3.6	2.5	0.0	13.2	4.7	2.8	0.0	14.4
CNG	10.9	3.0	4.2	20.6	12.7	3.7	3.0	21.5
EN	2.5	1.6	0.0	11.1	3.8	2.0	0.0	10.2
FI	6.9	3.1	0.0	15.3	9.3	3.6	2.1	20.4
HE	6.7	2.5	1.5	17.6	5.5	2.3	0.5	13.7
SE	4.5	2.0	0.1	11.0	3.7	2.0	0.0	11.1
TE	3.6	2.3	0.0	11.7	3.8	2.1	0.0	9.8
UT	39.4	5.6	24.9	53.6	34.3	5.5	20.9	50.3
L	16.8	3.9	6.3	29.3	15.0	4.3	3.5	29.1
M	27.5	5.3	14.0	43.5	25.1	4.7	12.3	39.0
S	55.7	5.4	41.0	72.7	59.9	5.9	42.9	78.2

Table 6: Summary statistics for the portfolio weights implied by SEARCH and CC without short sales. In sample: January 1998 - December 1999. Out of sample: January 2000 - June 2001. *m*: sample average. *s*: sample standard deviation. min, max: sample minimum and maximum.

In sample	CC		SEARCH		Out of sample	CC		SEARCH	
	W	N	W	N		W	N	W	N
<i>m</i>	0.007	0.003	-0.012	0.016		0.061	0.067	0.117	0.081
<i>s</i>	0.527	0.580	0.689	0.703		0.745	0.775	0.927	0.903
min	-2.020	-2.302	-2.338	-3.040		-2.435	-2.603	-2.880	-3.693
max	1.714	1.773	2.667	2.614		3.618	3.661	3.233	4.105

Table 7: Summary statistics for the returns of SEARCH and CC portfolios. W: with short sales. W: WSS, with short sales. N: NSS, no short sales. *m*: sample average. *s*: sample standard deviation. min, max: sample minimum and maximum.

In Sample	CC W	N	SEARCH W	N	Out of sample	CC W	N	SEARCH W	N
m	0.176	0.244	0.377	0.544		0.210	0.295	0.463	0.693
s	0.017	0.028	0.076	0.134		0.027	0.039	0.095	0.172
min	0.139	0.190	0.259	0.347		0.160	0.215	0.265	0.347
max	0.228	0.360	0.679	1.086		0.318	0.437	0.910	1.336

Table 8: Summary statistics for estimated and predicted volatilities of SEARCH and CC portfolios. W: WSS, with short sales. N: NSS, no short sales. m : sample average. s : sample standard deviation. min,max: sample minimum and maximum.

	Number of exceptions				Percentage of exceptions			
	CC		SEARCH		CC		SEARCH	
	W	N	W	N	W	N	W	N
Back-testing 1%	8	5	3	1	3.2%	2.0%	1.2%	0.4%
Forecasts 1%	27	19	15	7	7.2%	5.0%	4.0%	1.9%
Back-testing 5%	12	22	16	6	4.8%	8.8%	6.4%	2.4%
Forecasts 5%	31	50	36	22	8.2%	13.3%	9.5%	5.8%

Table 9: Exceptions to the value-at-risk calculated for the SEARCH and CC portfolios. W: WSS, with short sales. N: NSS, no short sales. Back testing on the last 250 observations in the estimation sample, forecast on 377 observations.

In sample	CC			SEARCH		
	G	W	N	G	W	N
m	1.54	1.57	1.37	1.19	1.23	1.9
s	2.46	2.51	2.19	1.78	1.96	1.55
min	2.5e-5	9.2e-7	1.9e-6	3.8e-6	2.9e-5	3.5e-6
max	20.41	21.64	18.68	17.7	17.14	16.43
Out of sample	CC			SEARCH		
	G	W	N	G	W	N
m	2.82	2.62	2.01	2.06	1.87	1.18
s	5.11	4.43	3.22	3.79	2.85	1.88
min	3.1e-6	1.1e-5	1.3e-6	3.6e-6	6.1e-7	1.3e-5
max	47.44	49.5	35.91	45.28	19.2	15.92

Table 10: Summary statistics for the EC_t ratio between realized portfolio variances and estimated portfolio variances by SEARCH and CC. G: Global minimum portfolio. W: WSS, with short sales. N: NSS, no short sales. m : sample average. s : sample standard deviation. min,max: sample minimum and maximum.